(19) JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: **05165494 A**

(43) Date of publication of application: **02.07.93**

(51) Int. Cl **G10L 3/00**
G10L 3/00
G10L 3/00

(21) Application number: **03330807**

(22) Date of filing: **13.12.91**

(71) Applicant: **OSAKA GAS CO LTD OKI ELECTRIC IND CO LTD**

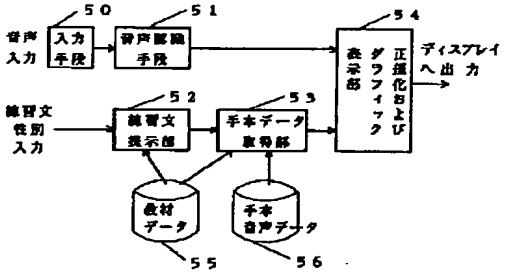
(72) Inventor: **OBA KATSUYA**
HIRAYAMA TERU
TAGAWA TADAMICHI
KATO MASAAKI

(54) VOICE RECOGNIZING DEVICE

(57) Abstract:

PURPOSE: To obtain a low-priced system facilitating the study of a learning person without being hourly limited by providing a correspondence display means to display the prescribed acoustic feature amount of a voice signal and a language symbol corresponding to the amount while making them correspondent.

CONSTITUTION: This device is equipped with an input means 50 to input the voice of an unspecified speaker, voice recognizing means 51 to obtain the language symbol by executing voice recognition according to the voice signal, and display means 54 to display the contents of making the prescribed acoustic characteristic amount of the voice signal correspondent to the language symbol corresponding to the amount by normalizing them concerning the two kinds of voices at least while making those amount and symbol correspondent. Namely, a normalizing part and graphic display part 54 inputs the voice data of the learning person outputted from the voice recognizing means 51 and model voice data from a model data acquisition part 52, normalizes speaking time for each word to which both data are correspondent, and graphically displays the voice data of the learning person and the model voice data (intonation, stress).



(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平5－165494

(43)公開日　平成５年(1993)７月２日

(51)Int.Cl.[5]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ１０Ｌ　3/00	５５１　Ｅ	8842－5H		
	Ｓ	8946－5H		
	５６１　Ｃ	8842－5H		

審査請求　未請求　請求項の数５(全 18 頁)

(21)出願番号　特願平3－330807

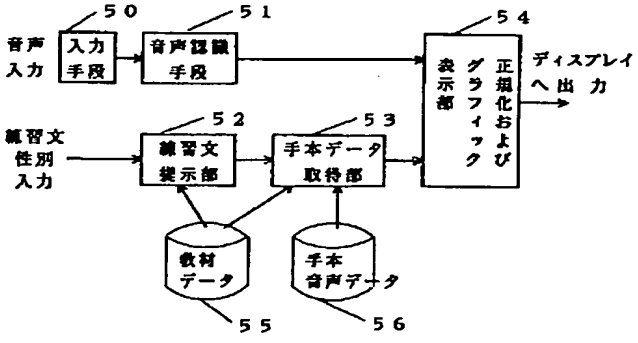
(22)出願日　平成３年(1991)12月13日

(71)出願人　000000284
大阪瓦斯株式会社
大阪府大阪市中央区平野町四丁目１番２号
(71)出願人　000000295
沖電気工業株式会社
東京都港区虎ノ門１丁目７番12号
(72)発明者　大場　克哉
大阪府豊中市新千里西町１丁目２番１号
株式会社オージー情報システム総研内
(72)発明者　平山　輝
大阪市中央区平野町四丁目１番２号　大阪瓦斯株式会社内
(74)代理人　弁理士　松田　正道

最終頁に続く

(54)【発明の名称】　音声認識装置

(57)【要約】
【目的】時間的制約を受けず、安価で、しかも学習者の学習が容易であるようなシステムを実現できる音声認識装置を提供すること。
【構成】不特定話者の音声を入力する入力手段５０と、その音声信号から音声認識を行い、言語シンボルを得る認識手段５１と、前記音声信号の所定の音響的特徴量とそれに対応する前記言語シンボルとの対応付け内容を、少なくとも２種類の音声について、正規化して対応させ表示する表示手段５４とを備えた音声認識装置である。



1

【特許請求の範囲】

【請求項１】　音声を入力する入力手段と、その音声信号から音声認識を行い、言語シンボルを得る認識手段と、前記音声信号の所定の音響的特徴量とそれに対応する前記言語シンボルとを対応付けながら表示する対応表示手段とを備えたことを特徴とする音声認識装置。

【請求項２】　不特定話者の音声を入力する入力手段と、その音声信号から音声認識を行い、言語シンボルを得る認識手段と、前記音声信号の所定の音響的特徴量とそれに対応する前記言語シンボルとの対応付け内容を、少なくとも２種類の音声について、正規化して対応させ表示する表示手段とを備えたことを特徴とする音声認識装置。

【請求項３】　不特定話者の音声を入力する入力手段と、その音声信号から音声認識を行い、言語シンボルを得る認識手段と、前記音声信号の所定の音響的特徴量とそれに対応する前記言語シンボルとの対応付け内容を、少なくとも２種類の音声について、比較する比較手段とを備えたことを特徴とする音声認識装置。

【請求項４】　前記２種類の音声は、手本となる音声と、学習者の音声であり、教育目的に用いられることを特徴とする請求項２又は３記載の音声認識装置。

【請求項５】　前記手本となる音声についての前記対応付け内容は、予め獲得され、記憶手段に格納されていることを特徴とする請求項４記載の音声認識装置。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、教育などの分野に応用可能な音声認識装置に関するものである。

【０００２】

【従来の技術】従来、例えば、外国語教育における発音（スピーキング）の訓練を行う方法としては次のようなものが知られている。（１）外国語教育専門家やネイティブスピーカーに自分の発音を聞いてもらい、アドバイスを受け、学習者の欠点を指摘してもらう。（２）市販の外国語学習用テープや２カ国語放送などを聞き、それをまねて発音する。また、参考書などに記載されている口の形や調音点等を示す図表をもとに練習する。自分の発音を、磁気テープなどに録音し、学習用テープに録音されている音声と聞き比べてみる。（３）計算機等を利用した発音練習装置により練習する。すなわち、発音練習用のＣＡＩ（computer assisted instruction：コンピュータ利用の教育）システムが知られている。これは、ネイティブスピーカーの発音を聞く。学習者の声を録音し、再生する。ネイティブスピーカーの音声波形と、学習者の音声波形をグラフィック表示する。また、ホルマントの表示や、時間毎の周波数や音圧の変化が表示される。

【０００３】

【発明が解決しようとする課題】しかしながら、上記の

2

ような従来の発音練習方法では次のような課題がある。（１）外国語教育専門家やネイティブスピーカーによる訓練方法では、大きな効果が期待できるが、時間的な制約を受け、また費用もかかる。（２）音声テープなどによる方法では、いつでも練習でき、また安価ではあるが、発した発音を客観的に評価できないため、学習者は自分の発音の何処が良くないかを把握することが困難である。（３）計算機などを利用して発音練習を行う方法では、確かにネイティブスピーカーの音声波形と学習者の音声波形が表示されるが、そこから何をどの様に読み取れば、学習者の発音がネイティブスピーカーの発音に近づくかが分かりにくいという課題がある。

【０００４】本発明は、このような従来の発音訓練方法の課題を考慮し、時間的制約を受けず、安価で、しかも学習者の学習が容易であるようなシステムを実現できる音声認識装置を提供することを目的とするものである。

【０００５】

【課題を解決するための手段】第１の本発明は、音声を入力する入力手段と、その音声信号から音声認識を行い、言語シンボルを得る認識手段と、前記音声信号の所定の音響的特徴量とそれに対応する前記言語シンボルとを対応付けながら表示する対応表示手段とを備えた音声認識装置である。

【０００６】第２の本発明は、不特定話者の音声を入力する入力手段と、その音声信号から音声認識を行い、言語シンボルを得る認識手段と、前記音声信号の所定の音響的特徴量とそれに対応する前記言語シンボルとの対応付け内容を、少なくとも２種類の音声について、正規化して対応させ表示する表示手段とを備えた音声認識装置である。

【０００７】第３の本発明は、不特定話者の音声を入力する入力手段と、その音声信号から音声認識を行い、言語シンボルを得る認識手段と、前記音声信号の所定の音響的特徴量とそれに対応する前記言語シンボルとの対応付け内容を、少なくとも２種類の音声について、比較する比較手段とを備えた音声認識装置である。

【０００８】

【作用】第１の本発明では、入力手段で入力された音声の信号から、認識手段によって音声認識を行い、言語シンボルを得る。また、対応表示手段は、音声信号の所定の音響的特徴量とそれに対応する言語シンボルとを対応付けながら表示する。

【０００９】第２の本発明では、入力手段で入力された不特定話者の音声信号から音声認識を行い、言語シンボルを得る。表示手段は、音声信号の所定の音響的特徴量とそれに対応する言語シンボルとの対応付け内容を、少なくとも２種類の音声について、正規化して表示する。

【００１０】第２の本発明では、入力手段で入力された不特定話者の音声信号から音声認識を行い、言語シンボルを得る。音声信号の所定の音響的特徴量とそれに対応

する言語シンボルとの対応付け内容を、比較手段によって、少なくとも２種類の音声について、比較する。

【００１１】

【実施例】以下本発明の一実施例について図面を参照しながら説明する。

【００１２】図１は本発明の音声認識装置の一実施例である音声データのグラフィック表示を行うための、コンピュータを利用したシステムを示す図である。同図に於て、学習者の音声がマイク１５から入力され、不特定話者連続音声認識手段１４により解析される。本発明の処理を行うプログラムと必要なデータが磁気ディスク１３に格納されており、ＣＰＵ１１により、メモリ１２にロードされる。キーボード１７、マウス１８は、後に述べる練習文を選択する際に使用され、イントネーション、ストレスなどがディスプレイ１６にグラフィック表示されるようになっている。

【００１３】図２は、本発明で使用する不特定話者連続音声認識手段の出力の一例である音声データである。認識した単語数２１、各単語の文字列の配列２２、各単語のセグメント（認識された音素）数の配列２３、各セグメントのピッチ周波数の平均の配列２４、各セグメントの音圧の平均の配列２５、各セグメントの発話時間の配列２６からなる。

【００１４】以下の実施例１は、第１の本発明の一実施例であって、認識対象である音声信号のイントネーションまたはストレスを、認識結果である単語に対応させて表示する装置を示す。

【００１５】また、以下の説明で配列のｎ番目とは、最初の要素を０番目とし、以下１番目、２番目．．．とする。

【００１６】（実施例１）図３は、本発明の実施例１のブロック図である。マイクロフォン１５等の入力手段３０から学習者の音声を入力し、図２に示した音声データを出力する、認識手段の一例としての音声認識処理部３１、および音声認識処理部３１から出力される音声データを入力とし、これから入力音声のイントネーション、ストレスを認識結果の単語列に対応させてディスプレイに表示する、対応表示手段の一例としてのグラフィック表示部３２からなる。

【００１７】図４は、その音声認識処理部３１の内容を示すフローチャートである。すなわち図４で、ステップ３１０では、マイクロフォン１５からの音声信号についてアナログ処理、Ａ／Ｄ変換がおこなわれる。ここで、入力：アナログ連続音声信号、出力：デジタル音声信号、処理概要：アンチ・エイリアスフィルタ、サンプリング周波数１６ｋＨｚで、デジタル信号に変換するものである。

【００１８】ステップ３１１では、音響パラメータ変換がおこなわれる。ここで、入力：デジタル音声信号、出力：フレーム（６．６ミリ秒）毎の２３の音響的特徴量、処理概要：デジタル信号を６．６ミリ秒のフレームに分け、スペクトル分析などにより、２３の音響的特徴量を抽出するものである。

【００１９】ステップ３１２では、音素コード（音素片）変換がおこなわれる。ここで、入力：フレーム毎の音響的特徴量、出力：セグメント毎の音素コード（音素片）、およびセグメント毎の音響的特徴量、処理概要：フレーム毎の音響的特徴量をもとに各フレームに音素コード（約１８００種類）を割り当てるものである。似通った特徴を持つ隣あったフレームはまとめ、１つのセグメントとするものである。

【００２０】ステップ３１３では、音素コード（音素片）／音素変換がおこなわれる。ここで、入力：セグメント毎の音素コード（音素片）およびセグメント毎の音響的特徴量、出力：セグメント毎の音素の候補、およびセグメント毎の音響的特徴量、外部参照データ：音素コードブック３１５、処理概要：音素コードブックを参照し、セグメント毎にそのセグメントが各音素（約５０種類）に対応する可能性を割り当てる。

【００２１】ステップ３１４では、音素／単語列変換がおこなわれる。ここで、入力：セグメント毎の音素の候補およびセグメント毎の音響的特徴量、出力：単語列およびセグメント毎の音響的特徴量、外部参照データ：単語辞書３１６、処理概要：単語辞書を参照し、入力のセグメント列と各単語に対応する音素列との距離を計算し、最も近いものを認識結果として出力する。

【００２２】なお、上記音素コードブック３１５は、音素コード（音素片）毎にその音素コードがある音素である可能性をもつテーブルであり、また、上記単語辞書３１６は、単語文字列とそれに対応する音素列を持つ辞書である。

【００２３】図５及び図６は、図３のグラフィック表示部３２の処理フローの例である。

【００２４】４０１は、前処理であり、ディスプレイにＸ軸、Ｙ軸を表示し、また、各変数の初期化を行う。すなわち、処理中の単語が認識結果の単語列の何番目かを示す変数L_word_cnt、処理中のセグメントが認識結果のセグメントの配列の何番目かを示す変数L_seg_cnt、グラフィック表示する折れ線グラフ（後述する図１５参照）の各標本点の値を示す変数（配列）の０番目L_x[0],L_y[0]にそれぞれ０を代入する。

【００２５】４０２では、音声認識結果の音声データの単語数をL_word_numに代入する。

【００２６】４０３では、L_word_cntとL_word_numの比較を行う。L_word_cntがL_word_numより小さければ、処理すべき単語が存在するということであり、ステップ４０４へ進む。L_word_cntがL_word_num以上であれば全ての単語についてグラフィック表示が終了したということであり、全ての処理を終える。

【００２７】４０４では、次の折れ線グラフ表示時にＸ

軸の下部に現在処理している単語を表示することを示すフラグdisp_word_flagをＯＮにする。

【００２８】４０５では、現在処理している単語が何番目のセグメントまでかを示す変数L_seg_numに音声データの各単語のセグメント数の配列のL_word_cnt番目の要素を加える。

【００２９】４０６では、L_seg_cntとL_seg_numとの比較を行う。L_seg_cntがL_seg_numより小さければ、現在処理している単語に対応するセグメントに未処理のセグメントが存在するということであり、ステップ４０７へ進む。L_seg_cntがL_seg_num以上であれば、現在処理している単語に対応する全てのセグメントに対し処理を終えたということであり、ステップ４１０へ進む。

【００３０】４０７では、表示すべき折れ線グラフの標本点を取得する。すなわち、L_x[L_seg_cnt + 1]に、{L_x[L_seg_cnt] + （音声データのセグメントの発話時間の配列のL_seg_cnt番目の要素)を代入し、L_y[L_seg_cnt + 1]に、（音声データのセグメントのピッチ周波数の平均の配列のL_seg_cnt番目の要素）または、（音声データのセグメントの音圧の平均の配列のL_seg_cnt番目の要素）を代入する。イントネーションの折れ線グラフ表示時は、セグメントのピッチ周波数の平均の配列の要素、ストレスの折れ線グラフ表示時には、セグメントの音圧の平均の配列の要素を代入する。

【００３１】４０８では、実際に折れ線グラフの表示を行う。すなわち、{(L_x[L_seg_cnt],L_y[L_seg_cnt),(L_x[L_seg_cnt + 1],L_y[L_seg_cnt +1])}の２点を結ぶ線分を表示する。

【００３２】４０９では、Ｘ軸の下部に対応する単語文字列を表示するかどうかの判断を行う。つまり、disp_word_flagがＯＮであるかどうか調べ、ＯＮであればステップ４１０へＯＦＦであれば、ステップ４１２へ進む。

【００３３】４１０では、(L_x[L_seg_cnt + 1],0)の直下に音声データの単語文字列の配列２２のL_word_cnt番目の要素である文字列を表示する。

【００３４】４１１では、１つの単語につき単語文字列の表示は１度でいいのでdisp_word_flagをＯＦＦにする。

【００３５】４１２では、処理を次のセグメントに移すため、L_seg_cntに１を加え、ステップ４０６へ戻る。

【００３６】４１３では、処理を次の単語に移すため、L_word_cntに１を加え、ステップ４０３へ戻る。

【００３７】次に、データの具体例を用いて上記の処理をより具体的に説明する。

【００３８】学習者のマイク１５を使用して音声を入力すると、音声認識手段３１でスペクトル分析等の手法により音声認識処理を行う。ここでは、音声認識手段３１の出力として、図２に示した音声データが得られたとする。

【００３９】グラフィック表示部３２では、この音声データを入力として以下の処理を行う。

【００４０】まず、前処理として、Ｘ軸、Ｙ軸の表示、L_word_cnt、L_seg_cnt、L_x[0]、L_y[0]にそれぞれ０を代入する（ステップ４０１）。続いて、L_word_numに音声データの単語数の値である４を代入する（ステップ４０２）。

【００４１】ここで、L_word_cntとL_word_numの比較を行う（ステップ４０３）が、今L_word_cntが０、L_word_numが４なので、ステップ４０４へ進む。

【００４２】ステップ４０４では、今から処理する単語の文字列を次の折れ線表示時にＸ軸の下部に表示することを示すdisp_word_flagをＯＮにする。次に、L_seg_numに音声データの各単語のセグメント数の配列のL_word_cnt番目の要素を加える。ここで、L_seg_num＝０、L_word_cnt＝０である。音声データの各単語のセグメント数の配列の０番目の要素は３であるのでL_seg_num＝３となる（ステップ４０５）。

【００４３】ステップ４０６では、L_seg_cntとL_seg_numの比較を行う。今、L_seg_cnt＝０、L_seg_num＝３なので、処理はステップ４０７へ進む。

【００４４】ここで、実際に折れ線グラフ表示を行うための標本点を取得する。すなわち、L_x[1]にL_x[0]の値と音声データの各セグメントの発話時間の配列の０番目の要素の値を加え、２４を代入する。また、例えばイントネーションのグラフを表示するのであれば、L_y[1]に音声データの各セグメントのピッチ周波数の平均の配列の０番目の要素の値１０４を代入する（ステップ４０７）。続いて実際に、折れ線グラフを表示する。つまり、{(L_x[0], L_y[0]), (L_x[1], L_y[1])}の値である{(0,0),(24,104)}の２点を結ぶ線分を表示する（ステップ４０８）。

【００４５】次に、disp_word_flagがＯＮであるかどうかを調べ（ステップ４０９）、今ＯＮなので、Ｘ軸の下部に対応する単語文字列を表示するための処理を行う。つまり、音声データの単語文字列の配列の０番目の値である"how"を(x[1],0)の直下に表示する（ステップ４１０）。"how"を処理している間、もうこの文字列を表示する必要はないので、disp_word_flagをＯＦＦにする（ステップ４１１）。

【００４６】０番目のセグメントに対する処理が終了したので、L_seg_cntに１を加え１として、ステップ４０６へ戻る。１番目のセグメントについても上記と同様に折れ線グラフ表示を行う（ステップ４０７、４０８）。

【００４７】ステップ４０９では、disp_word_flagがＯＦＦになっているので、単語文字列の表示処理は行わず、ステップ４１２へ進む。

【００４８】このようにして、２番目のセグメントまでの処理が終わったとする。このとき、L_seg_cntが３となっている（ステップ４１２）。ステップ４０６での比較の結果、L_seg_cntがL_seg_num以上になったので、ス

テップ４１３へ進み、次の単語へ処理を進めるために、L_word_cntに１を加え、ステップ４０３へ戻る。

【００４９】同様にして、４番目の単語まで全ての処理が終わったとする。すると、L_word_cntが４となり（ステップ４１３）、４０３での比較の結果全ての単語に付いて処理が終わったとして、本実施例の処理が終了する。

【００５０】（実施例２）実施例２は、第２の本発明の音声認識装置の一実施例であって、外国語発音練習において、手本となるある文章を入力した音声のイントネーションまたはストレスを、認識した単語に対応させてグラフィック表示し、さらに学習者の同じ文章の音声のイントネーションまたはストレスを、認識した単語を一致させ正規化して、単語毎に、手本のグラフィックに対応させて表示する装置の例である。

【００５１】図７は、その実施例２のブロック図である。

【００５２】音声認識手段５１は、学習者が音声を入力し、認識結果として、音声データを出力する手段であり、実施例１の音声認識手段３１と同一の機能を持つ。また、音声認識処理の結果得られる単語列は、無音部分を除くと、学習者により選択された練習文の練習文文字列６１（図８参照）と同じであるとする。

【００５３】練習文提示部５２は、教材データ５５に保存されている教材ファイルの練習文番号６０と練習文文字列６１を利用して練習文のリストを表示し、学習者にこれから練習する文と自分の性別の入力を促す。学習者により練習文が選択され、性別が入力されると、選択された練習文に対応する練習文番号と性別を出力する。

【００５４】手本データ取得部５３は、練習文提示部５２の出力である学習者が選択した練習文の番号と学習者の性別を入力とし、教材データ５５を読み込み、練習文の番号に対応する音声データファイルを手本音声データ５６から得る手段である。

【００５５】本発明の表示手段の一例としての正規化部およびグラフィック表示部５４は、音声認識手段５１から出力される学習者の音声データと手本データ取得部５３からの手本音声データを入力とし、両者の対応する単語毎の発話時間を正規化し、学習者の音声データと手本の音声データ（イントネーション、ストレス）をグラフィック表示する手段である。図１５及び図１６は、そのようなグラフィック表示手段５４の画面の一例である。

【００５６】図８は、教材ファイルの例であり、練習文番号６１、練習文の文字列６２、男性用の音声データファイル名６３、女性用の音声データファイル姪４からなる。

【００５７】図９は、手本音声データファイルの例であり、学習者の音声を認識した結果の音声データと同じ構造の手本音声データが保存されている。これら手本のデータは、予め第１の本発明で述べたようにして得られたものである。

【００５８】図１０は、正規化及びグラフィック表示部５４の処理フローの例である。

【００５９】８０１は、前処理であり、Ｘ軸、Ｙ軸の表示、現在処理している学習者の音声データの単語が何番目かを示すL_word_cnt、現在処理している学習者の音声データのセグメントが何番目かを示すL_seg_cnt、学習者のイントネーション（ストレス）の折れ線グラフ表示時の標本点となる配列の０番目の要素L_x[0]，L_y[0]，現在処理している手本音声データの単語が何番目かを示すM_word_cnt，現在処理している手本音声データのセグメントが何番目かを示すM_seg_cnt，手本のイントネーション（ストレス）の折れ線グラフ表示時の標本点となる配列の０番目の要素M_x[0]，M_y[0]のそれぞれに０を代入する。

【００６０】８０１１では、音声認識結果の音声データの単語数をL_word_numに代入する。

【００６１】８０２は、全ての単語について、処理が終わったかどうかを調べるために、L_word_cntとL_word_numの比較を行う。L_word_cntがL_word_numより小さいとは、まだ処理すべき単語が残っているということであり、ステップ８０３へ進み、そうでないときは、全ての処理を終了する。

【００６２】８０３では、手本音声データのM_word_cnt番目の単語について折れ線グラフと対応する単語文字列の表示を行う。このステップの詳細なフローの例を図１１及び図１２に示す。

【００６３】８０５では、L_word_cnt番目の学習者の単語に対応するイントネーションまたはストレスを手本のイントネーションまたはストレスに対応させて表示するための倍率を求める。この処理の詳細フローの例を図１３に示す。

【００６４】８０６では、L_word_cnt番目の学習者の単語に対応するイントネーションまたはストレスを折れ線グラフ表示する。この時８０５で求めた倍率により手本の対応する単語と同じ位置（Ｘ座標）に表示する。

【００６５】８０７では、学習者、手本共に次の単語へ処理を進めるため、L_word_cnt，M_word_cntの値にそれぞれ１を加える。

【００６６】図１１及び図１２は、図１０のステップ８０３の詳細フローの例である。

【００６７】９０１では、次の折れ線グラフ表示時にＸ軸の下部に単語文字列を表示することを示すdisp_word_flagをＯＮにする。また、この単語を発話するのに要した時間を示す変数M_word_durに０を代入する。

【００６８】９０２では、現在処理している単語が何番目のセグメントまでかを示す変数M_seg_numに音声データの各単語のセグメント数の配列のM_word_cnt番目の要素を加える。

【００６９】９０３では、M_seg_cntとM_seg_numとの比

較を行う。M_seg_cntがM_seg_numより小さければ、現在処理している単語に対応するセグメントに未処理のセグメントが存在するということであり、ステップ９０４へ進む。M_seg_cntがM_seg_num以上であれば、現在処理している単語に対応する全てのセグメントに対し処理を終えたということであり、処理を終了する。

【００７０】９０４では、表示すべき折れ線グラフの標本点を取得する。すなわち、M_x[M_seg_cnt + 1]に、{M_x[M_seg_cnt] + (音声データのセグメントの発話時間の配列のM_seg_cnt番目の要素}を代入し、M_y[M_seg_cnt + 1]に、(音声データのセグメントのピッチ周波数の平均の配列のM_seg_cnt番目の要素)または、(音声データのセグメントの音圧の平均の配列のM_seg_cnt番目の要素）を代入する。イントネーションの折れ線グラフ表示時は、セグメントのピッチ周波数の平均の配列の要素、ストレスの折れ線グラフ表示時には、セグメントの音圧の平均の配列の要素を代入する。

【００７１】また、同時にM_word_durに音声データのセグメントの発話時間の配列のM_seg_cnt番目の要素を加える。

【００７２】９０５では、実際に折れ線グラフの表示を行う。すなわち、{(M_x[M_seg_cnt],M_y[M_seg_cnt),(M_x[M_seg_cnt + 1],M_y[M_seg_cnt + 1])}の２点を結ぶ線分を表示する。

【００７３】９０６では、Ｘ軸の下部に対応する単語文字列を表示するかどうかの判断を行う。つまり、disp_word_flagがＯＮであるかどうか調べ、ＯＮであればステップ９０７へＯＦＦであれば、ステップ９０９へ進む。

【００７４】９０７では、(M_x[M_seg_cnt],0)の直下に認識結果の単語文字列の配列２２のM_word_cnt番目の要素である文字列を表示する。

【００７５】９０８では、１つの単語につき単語文字列の表示は１度でいいのでdisp_word_flagをＯＦＦにする。

【００７６】９０９では、処理を次のセグメントに移すため、M_seg_cntに１を加え、ステップ９０３へ戻る。

【００７７】図１３は、図１０のステップ８０５の詳細フローの例である。

【００７８】１００１では、現在処理している学習者の単語のセグメント数を示す変数L_word_seg_numに音声データの各単語のセグメント数の配列のL_word_cnt番目の値を代入し、また、この単語を発話するのに要した時間を示す変数L_word_dur,ループ用の変数ｉをそれぞれ０に初期化する。

【００７９】１００２では、ｉとL_word_seg_numの比較を行う。ｉがL_word_seg_numより小さければステップ１００３へ進み、ｉがL_word_seg_num以上であればステップ１００５へ進む。

【００８０】１００３では、現在処理している単語のｉ番目のセグメントの時間をL_word_durに加える。すなわち、学習者の音声データの各セグメントの発話時間の配列の(L_seg_cnt+i)番目の値をL_word_durに加える。

【００８１】１００４では、次のセグメントの発話時間を求めるためｉに１を加える。

【００８２】１００５では、現在処理している学習者の単語を表示するときの倍率を求め、処理を終了する。つまり、倍率を示す変数rateに(M_word_dur/L_word_dur)を代入し、処理を終える。

【００８３】図１４は、図１０のステップ８０６の詳細フローの例である。

【００８４】１１０１では、現在処理中の単語に対応するセグメントが全体で何番目のセグメントまでかを示す変数L_seg_numに音声データの各単語のセグメント数の配列のL_word_cnt番目の要素の値を加える。

【００８５】１１０２では、現在処理を行っている単語に対応する全てのセグメントに対し処理を終えたかどうかを調べる。すなわち、L_seg_cntとL_seg_numの値を比較し、L_seg_cntがL_seg_numより小さければステップ１１０３へ進み、そうでなければ処理を終了する。

【００８６】１１０３では、表示すべき折れ線グラフの標本点を取得する。すなわち、L_x[L_seg_cnt + 1]に、{L_x[L_seg_cnt] + (音声データのセグメントの発話時間の配列のL_seg_cnt番目の要素)*rate}を代入し、L_y[L_seg_cnt + 1]に、(音声データのセグメントのピッチ周波数の平均の配列のL_seg_cnt番目の要素)または、(音声データのセグメントの音圧の平均の配列のL_seg_cnt番目の要素）を代入する。イントネーションの折れ線グラフ表示時は、セグメントのピッチ周波数の平均の配列の要素、ストレスの折れ線グラフ表示時には、セグメントの音圧の平均の配列の要素を代入する。

【００８７】１１０４では、実際に折れ線グラフの表示を行う。すなわち、{(L_x[L_seg_cnt],L_y[L_seg_cnt),(L_x[L_seg_cnt + 1],L_y[L_seg_cnt + 1])}の２点を結ぶ線分を表示する。

【００８８】１１０５では、処理を次のセグメントに移すため、M_seg_cntに１を加え、ステップ１１０２へ戻る。

【００８９】次に、データの具体例を用いて上記の処理を説明する。

【００９０】練習文提示部５２により、図８の教材ファイルをもとに練習文リストが学習者に提示されたとする。

【００９１】続いて、学習者により、練習文として、"how do you do"が選択され、また、学習者は男性であるとする。

【００９２】手本データ取得部５３は、教材ファイル５５を参照し、"how do you do"の男性用の手本音声データファイル名"M_m1.dat"を得る。"M_m1.dat"の内容は、図９に示す手本データファイルの例と同一であるとする。

【００９３】さらに、学習者の音声入力による音声認識手段５１の出力は、図２に示すとおりであるとする。

【００９４】正規化及びグラフィック表示部５４では、学習者音声データ、手本音声データを入力とし、以下のように処理を行う。

【００９５】まず、Ｘ軸、Ｙ軸の表示、各種変数の初期化を行う（ステップ８０１）。

【００９６】次に、L_word_numに学習者の音声データの単語数を代入する。ここでは、４が代入される（ステップ８０１１）。

【００９７】ここで、L_word_cntとL_word_numの比較を行う（ステップ８０２）が、今L_word_cntが０、L_word_numが４なので、ステップ８０３へ進む。

【００９８】ステップ８０３では、L_word_cnt番目の手本音声データの単語に対応するイントネーション（ストレス）を表示する。

【００９９】すなわち、ステップ９０１では、手本におけるその単語の発話時間を示す変数M_word_durを０に初期化し、さらに今から処理する単語の文字列を次の折れ線グラフ表示時にＸ軸の下部に表示することを示すdisp_word_flagをＯＮにする。

【０１００】次に、M_seg_numに音声データの各単語のセグメント数の配列のL_word_cnt番目の要素を加える。ここで、M_seg_num＝０、L_word_cnt＝０である。音声データの各単語のセグメント数の配列の０番目の要素は２であるのでM_seg_num＝２となる（ステップ９０２）。

【０１０１】ステップ９０３では、M_seg_cntとM_seg_numの比較を行う。今、M_seg_cnt＝０、M_seg_num＝２なので、処理はステップ９０４へ進む。

【０１０２】ここで、実際に折れ線グラフ表示を行うための標本点を取得する。すなわち、M_x[1]にM_x[0]の値と音声データの各セグメントの発話時間の配列のM_seg_cnt＝０番目の要素の値を加え、１４を代入する。また、例えばイントネーションのグラフを表示するのであれば、M_y[1]に音声データの各セグメントのピッチ周波数の平均の配列の０番目の要素の値１１４を代入する。さらに、M_word_durに現在のM_word_durに手本音声データの各セグメントの発話時間の配列のM_seg_cnt＝０番目の値１４を加え、１４とする。（ステップ９０４）。

【０１０３】続いて実際に、折れ線グラフを表示する。つまり、{(L_x[0], L_y[0]), (L_x[1], L_y[1])}の値である{(0,0), (14,114)}の２点を結ぶ線分を表示する（ステップ９０５）。

【０１０４】次に、disp_word_flagがＯＮであるかどうかを調べ（ステップ９０６）、今ＯＮなので、Ｘ軸の下部に対応する単語文字列を表示するための処理を行う。つまり、音声データの単語文字列の配列の０番目の値である"how"を(x[1],0)の直下に表示する（ステップ９０７）。"how"を処理している間、もうこの文字列を表示

する必要はないので、disp_word_flagをＯＦＦにする（ステップ９０８）。

【０１０５】０番目のセグメントに対する処理が終了したので、M_seg_cntに１を加え１として（ステップ９０９）、ステップ９０３へ戻る。１番目のセグメントについても上記と同様に折れ線グラフ表示を行う（ステップ９０５、９０６）。

【０１０６】ステップ４０９では、disp_word_flagがＯＦＦになっているので、単語文字列の表示処理は行わず、ステップ９０９へ進む。

【０１０７】このとき、M_seg_cntが２となっている（ステップ９０９）。ステップ９０３での比較の結果、M_seg_cntがM_seg_num以上になったので、手本の０番目の単語に対する処理を終え、ステップ８０５へ進む。ここで、M_word_dur＝２７となっている。

【０１０８】次に、学習者のL_word_cnt番目の単語に対応するセグメントのデータを手本に対応させて表示するための倍率を計算する（ステップ８０５）。

【０１０９】すなわち、学習者のL_word_cnt番目の単語に対応するセグメント数を示す変数L_word_seg_numに音声データの各単語のセグメント数の配列のL_word_cnt＝０番目の値３を代入し、L_word_dur, ｉにそれぞれ０を代入する。

【０１１０】次に、L_word_cnt＝０番目の単語に対応する全てのセグメントに対して処理を終えたかどうか調べるために、ｉとL_word_seg_numを比較する。今、ｉ＝０、L_word_seg_num＝３なので、処理はステップ１００３へ進む。

【０１１１】ここで、L_word_durに現在のL_word_dur＝０と音声データの各セグメントの発話時間の配列の(L_seg_cnt + i)＝０番目の値２４の和である２４を代入する（ステップ１００３）。

【０１１２】さらに、次のセグメントの発話時間を調べるためｉに１を加える（ステップ１００４）。

【０１１３】このようにして、処理を２番目のセグメントまで終えたとする。このとき、ｉ＝３，L_word_dur＝４６となり、ステップ１００２の比較により、ステップ１００５へ進む。

【０１１４】ステップ１００５では、(M_word_dur＝２７)／(L_word_dur＝４６)をrateに代入する。すなわち、rateに約０．５９が代入される。

【０１１５】学習者の音声データを表示する際の倍率が求められた（ステップ８０５）ので、この倍率を利用して、L_word_cnt＝０番目の学習者の単語に対応するイントネーション（ストレス）のグラフィック表示を行う（ステップ８０６）。

【０１１６】ステップ１１０１では、L_seg_numに音声データの各単語のセグメント数の配列のL_word_cnt番目の要素を加える。ここで、L_seg_num＝０、L_word_cnt＝０である。音声データの各単語のセグメント数の配列

の０番目の要素は３であるのでL_seg_num＝３となる。

【０１１７】ステップ１１０２では、L_seg_cntとL_seg_numの比較を行う。今、L_seg_cnt＝０、L_seg_num＝３なので、処理はステップ１１０３へ進む。

【０１１８】ここで、実際に折れ線グラフ表示を行うための標本点を取得する。すなわち、L_x[1]にL_x[0]の値に音声データの各セグメントの発話時間の配列の０番目の要素の値と倍率rateの積を加える。ここでは、0+24*0.59=14.16を代入する。また、例えばイントネーションのグラフを表示するのであれば、L_y[1]に音声データの各セグメントのピッチ周波数の平均の配列の０番目の要素の値１０４を代入する（ステップ１１０３）。続いて実際に、折れ線グラフを表示する。つまり、{(L_x[0], L_y[0]), (L_x[1], L_y[1])}の値である{(0,0), (14.16,104)}の２点を結ぶ線分を表示する（ステップ１１０４）。

【０１１９】学習者のL_seg_cnt＝０番目のセグメントのデータの表示が終わったので、処理を次のセグメントに進めるため、L_seg_cntに１を加える（ステップ１１０５）。

【０１２０】このようにして、処理を進め、学習者の２番目のセグメントまでの処理が終わるとL_seg_cnt＝３となり、ステップ１１０２の比較の結果、学習者のL_word_cnt＝０番目の単語の処理を終わり、ステップ８０７へ進む。

【０１２１】ステップ８０７では、処理を次の単語に進めるため、L_word_cntに１を加え、１とし、ステップ８０２へ戻る。

【０１２２】このようにして処理を進め、３番目の単語まで処理を終えたとき、L_word_cnt＝４となり、ステップ８０２の比較により、本実施例の全ての処理を終了する。

【０１２３】なお、第３の本発明として、手本となる人の音響的特徴量と言語シンボルとの対応付け内容と、学習者のそれとを、上記実施例のように、正規化して表示するのではなく、あるいは表示に加えて、両者の対応付け内容を比較手段で比較して、その不一致の部分について、ある単語のイントネーションを下げるようになどという指示を表示するようにしてもよい。図１６の１３０はその指示の一例である。

【０１２４】また、手本となる音声データなどは、予め記憶手段に格納されてなく、いつでもネイティブスピーカーなどによって入力でき、音声信号の所定の音響的特徴量とそれに対応する言語シンボルとの対応付け内容が得られるようになっていてもよい。

【０１２５】また、本発明の音響的特徴量とは、イントネーションやストレスに限らず、フォルマント等の他の音響的特徴量であってもよい。

【０１２６】また、本発明の言語的シンボルとは、言葉、発音記号、音素コード、音素片など、言語に関するシンボルであればどのようなものでもよい。

【０１２７】また、本発明の各手段は、コンピュータを用いてソフトウェア的に実現しても、それら機能を有する専用のハード回路を用いて実現してもかまわない。

【０１２８】

【発明の効果】以上の説明から明らかなように、第１の本発明は、音声信号の所定の音響的特徴量とそれに対応する言語シンボルとを対応付けながら表示する対応表示手段を備えるので、その音響的特徴量と言語シンボルとの対応関係が分かりやすいという長所を有する。

【０１２９】また、第２の本発明は、音声信号の所定の音響的特徴量とそれに対応する言語シンボルとを対応付ける対応手段と、少なくとも２種類の音声について、対応手段によって得られた対応付け内容を比較する比較手段とを備えるので、例えば、１種類の音声をネイティブスピーカーの音声とし、他方の種類の音声を学習者の音声とすると、それらの比較によって、学習者の悪い所などを的確に指摘することが可能になる。また、第３の本発明は、音声信号の所定の音響的特徴量とそれに対応する言語シンボルとを対応付ける対応手段と、少なくとも２種類の音声について、各対応手段によって得られた対応付け内容を正規化して表示する表示手段とを備えるので、例えば、１種類の音声をネイティブスピーカーの音声とし、他方の種類の音声を学習者の音声とすると、それらの対応付け内容が正規化されて表示されるので、その表示を見て、学習者の悪い所などを的確に理解することが可能になる。

【図面の簡単な説明】

【図１】本発明の音声認識装置の一実施例を示すブロック図である。

【図２】本発明の音声認識手段の音声データである。

【図３】第１の本発明の音声認識装置の一実施例を示すブロック図である。

【図４】本発明の音声認識手段の動作を示すフローチャートである。

【図５】第１の本発明の対応表示手段の動作を説明するためのフローチャートである。

【図６】第１の本発明の対応表示手段の動作を説明するためのフローチャートである。

【図７】第２の本発明の音声認識装置の一実施例を示すブロック図である。

【図８】同実施例の教材ファイルの一例を示す構成図である。

【図９】同実施例の手本音声データファイルの一例を示す構成図である。

【図１０】同実施例の動作を説明するためのフローチャートである。

【図１１】図１０のステップ８０３の詳細フローチャートである。

【図１２】図１０のステップ８０３の詳細フローチャー

15

トである。

【図１３】図１０のステップ８０５の詳細フローチャートである。

【図１４】図１０のステップ８０６の詳細フローチャートである。

【図１５】第２の本発明の音声認識装置の表示画面を示す図である。

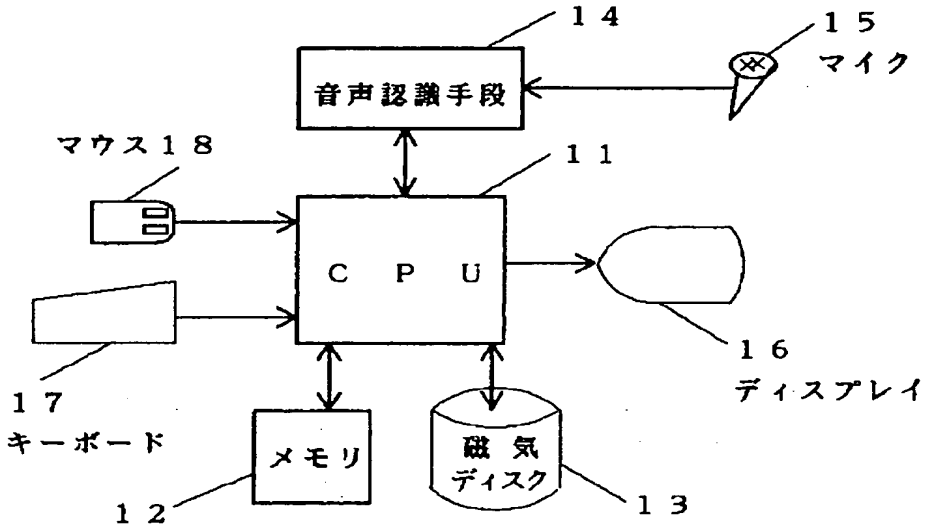
【図１６】第２の本発明の音声認識装置の表示画面を示す図である。

【符号の説明】

16

３０	入力手段
３１	音声認識手段
３２	グラフィック表示部
５０	入力手段
５１	音声認識手段
５２	練習文提示部
５３	手本データ取得部
５４	正規化及びグラフィック表示部
５５	教材データ
５６	手本音声データ

【図１】



【図２】

２１
単語数　　４

２２
単語文字列　　[how][do][you][do]

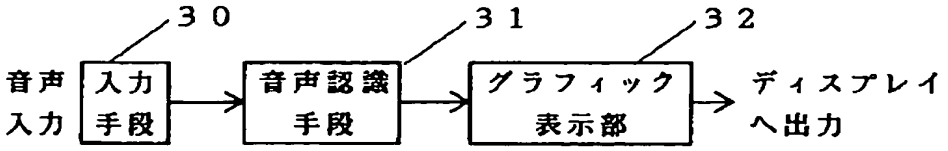
２３
各単語の　　[3][2][2][3]
セグメント数

２４
各セグメント　　[104][100][102][110][130][100][100][106][200][150]
のピッチ

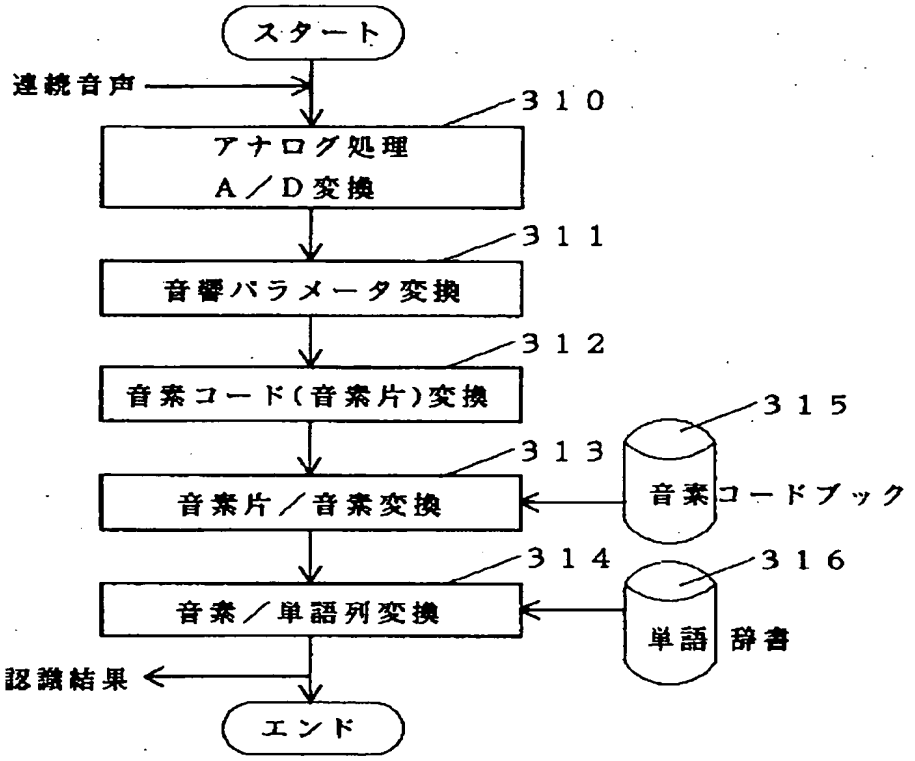
２５
各セグメント　　[154][200][12][140][150][140][200][126][230][57]
の音圧

２６
各セグメント　　[24][10][12][11][13][10][10][18][20][15]
の発話時間

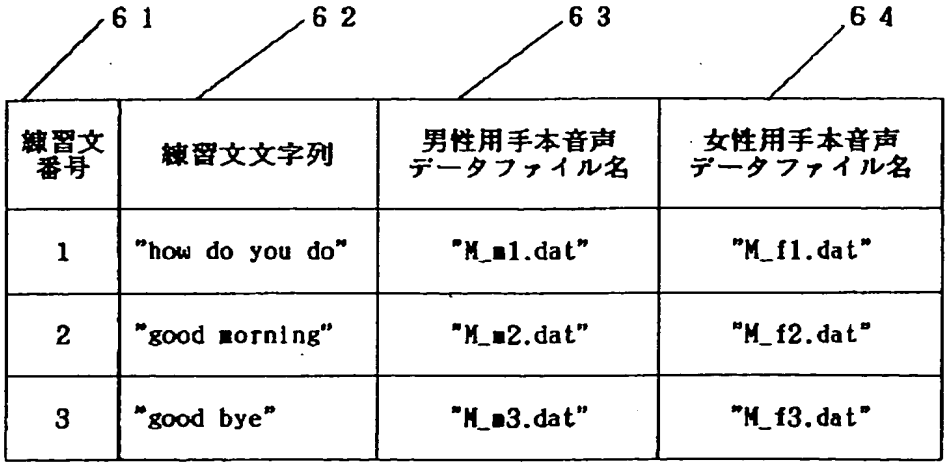
【図３】



【図４】

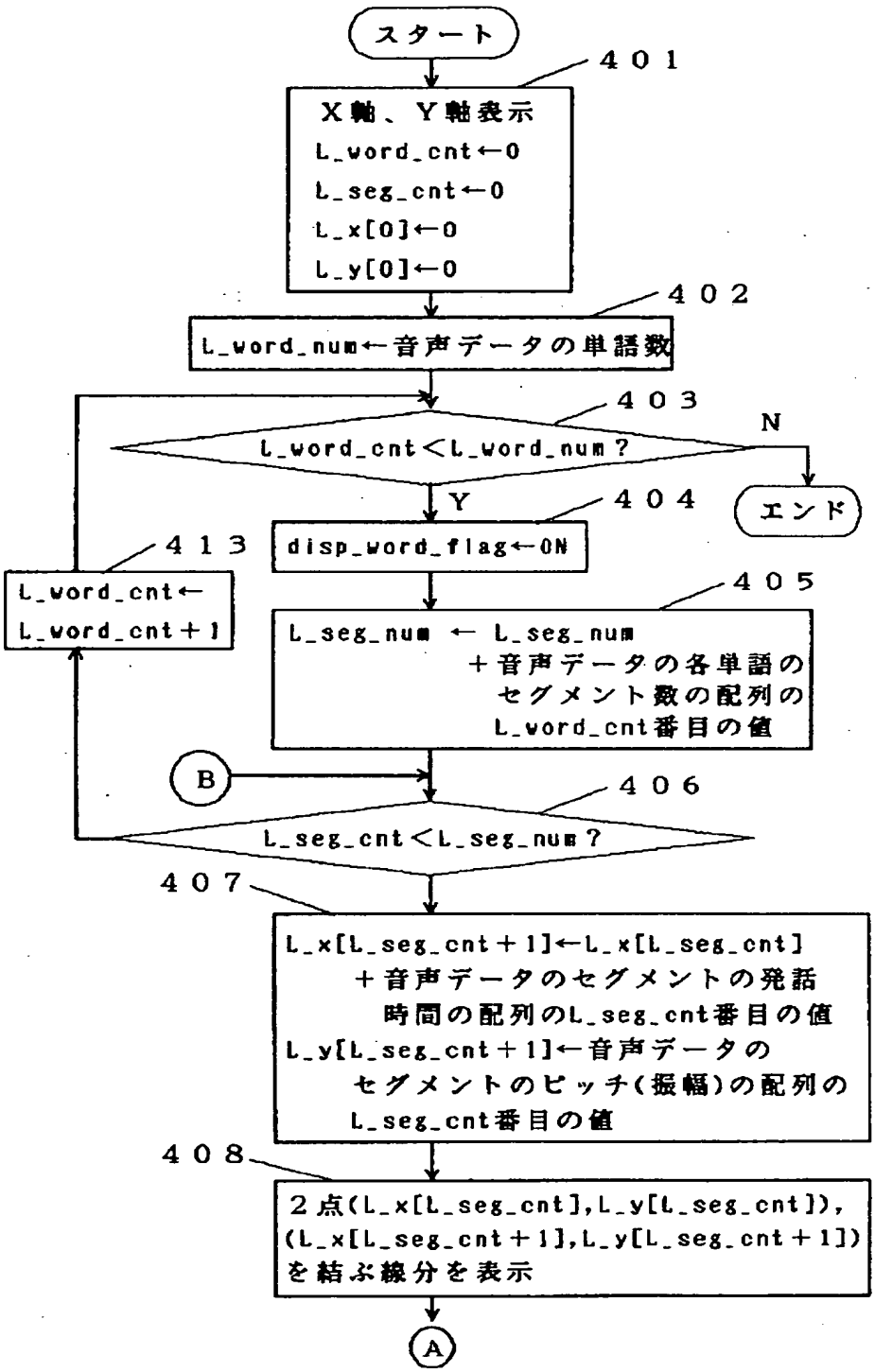


【図８】

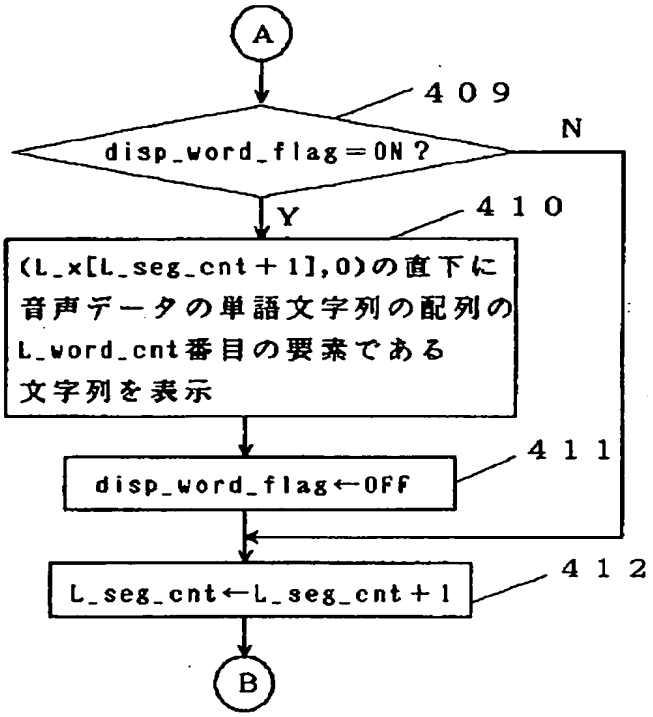


練習文番号	練習文文字列	男性用手本音声データファイル名	女性用手本音声データファイル名
1	"how do you do"	"M_m1.dat"	"M_f1.dat"
2	"good morning"	"M_m2.dat"	"M_f2.dat"
3	"good bye"	"M_m3.dat"	"M_f3.dat"

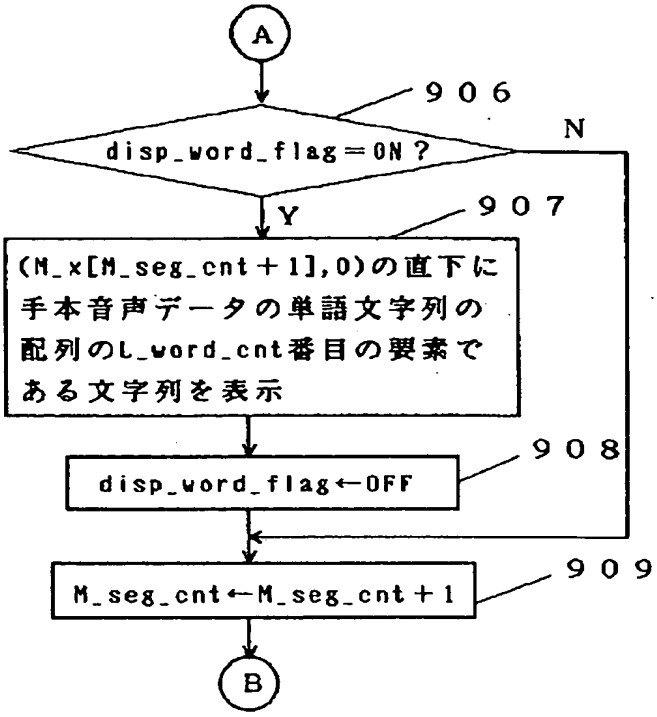
【図5】


スタート
401
X軸、Y軸表示
L_word_cnt←0
L_seg_cnt←0
L_x[0]←0
L_y[0]←0
402
L_word_num←音声データの単語数
403
L_word_cnt＜L_word_num？
N
エンド
Y
404
disp_word_flag←ON
413
L_word_cnt←
L_word_cnt＋1
405
L_seg_num ← L_seg_num
＋音声データの各単語の
セグメント数の配列の
L_word_cnt番目の値
B
406
L_seg_cnt＜L_seg_num？
407
L_x[L_seg_cnt＋1]←L_x[L_seg_cnt]
＋音声データのセグメントの発話
時間の配列のL_seg_cnt番目の値
L_y[L_seg_cnt＋1]←音声データの
セグメントのピッチ(振幅)の配列の
L_seg_cnt番目の値
408
2点(L_x[L_seg_cnt],L_y[L_seg_cnt]),
(L_x[L_seg_cnt＋1],L_y[L_seg_cnt＋1])
を結ぶ線分を表示
A

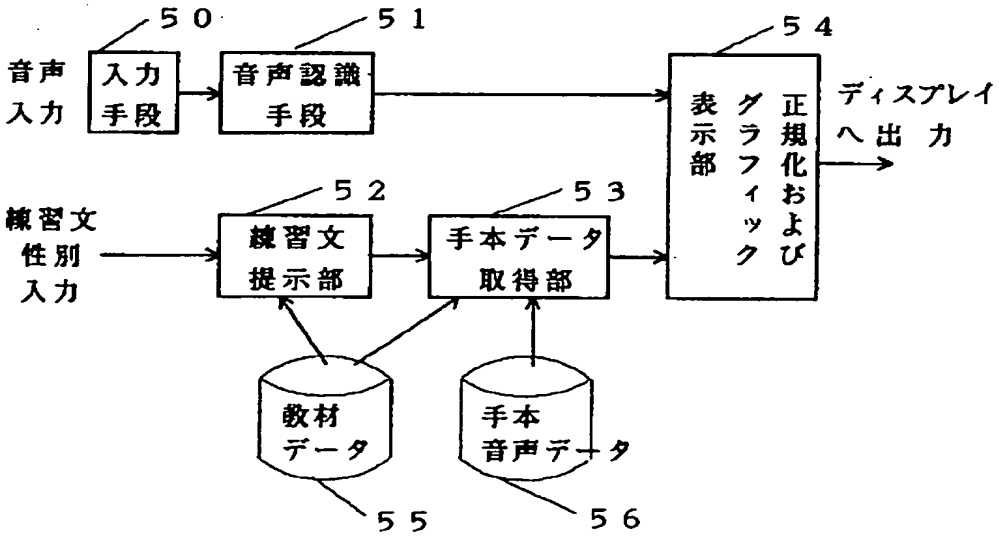
【図６】


A
４０９
disp_word_flag＝ON？
N
Y
４１０
(L_x[L_seg_cnt＋1],0)の直下に音声データの単語文字列の配列のL_word_cnt番目の要素である文字列を表示
４１１
disp_word_flag←OFF
４１２
L_seg_cnt←L_seg_cnt＋1
B

【図１２】


A
９０６
disp_word_flag＝ON？
N
Y
９０７
(M_x[M_seg_cnt＋1],0)の直下に手本音声データの単語文字列の配列のL_word_cnt番目の要素である文字列を表示
９０８
disp_word_flag←OFF
９０９
M_seg_cnt←M_seg_cnt＋1
B

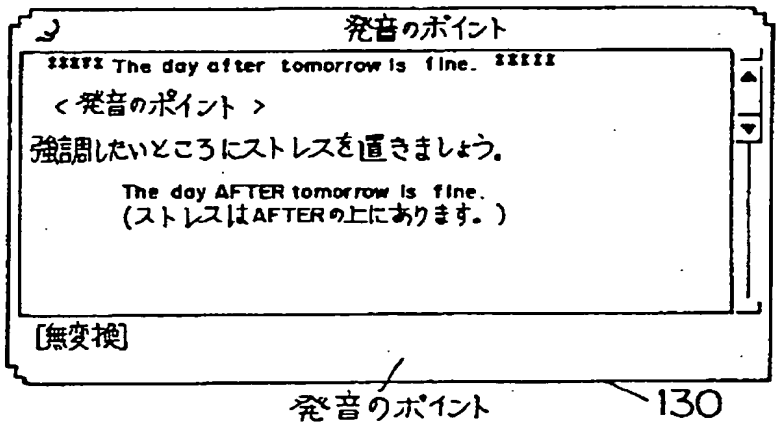
【図７】


５０
５１
５４
音声入力
入力手段
音声認識手段
正規化およびグラフィック表示部
ディスプレイへ出力
５２
５３
練習文性別入力
練習文提示部
手本データ取得部
教材データ
手本音声データ
５５
５６

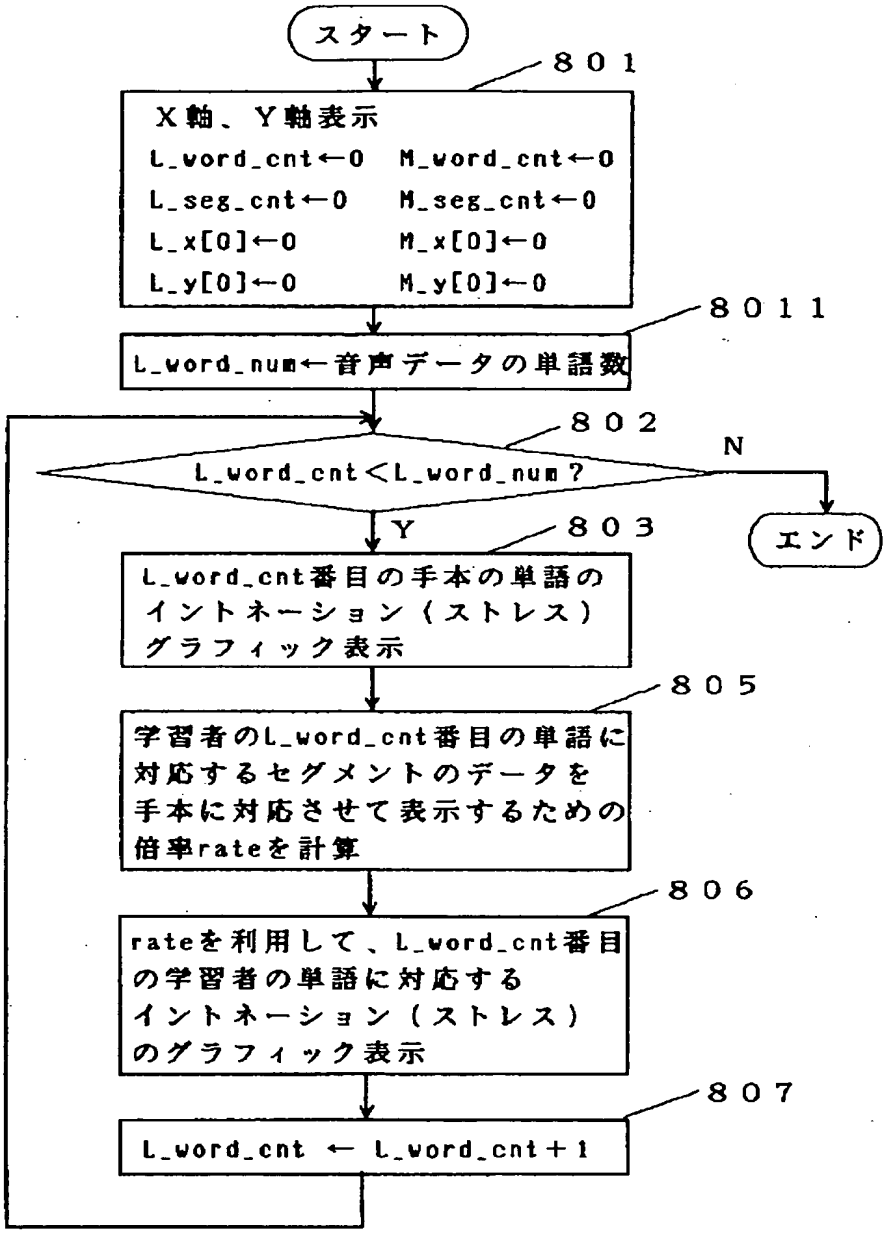
【図９】

単語数 71	4	
単語文字列 72	[hov][do][you][do]	
各単語の セグメント数 73	[2][1][2][3]	
各セグメント のピッチ 74	[114][100]	[100][100][98][106][200][150]
各セグメント の音圧 75	[144][100]	[110][208][158][86][130][97]
各セグメント の発話時間 76	[14][13]	[10][10][9][15][13][14]

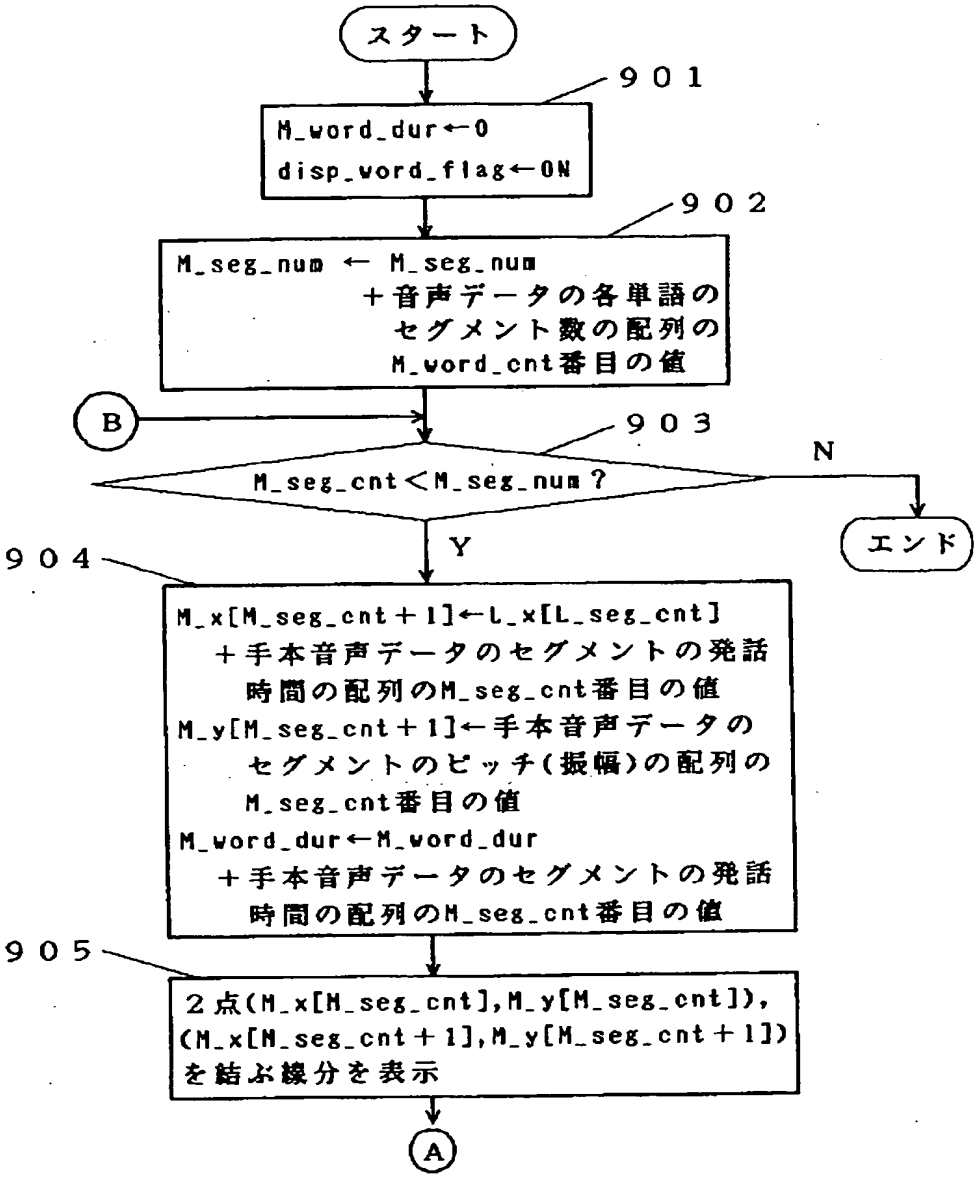
【図１６】



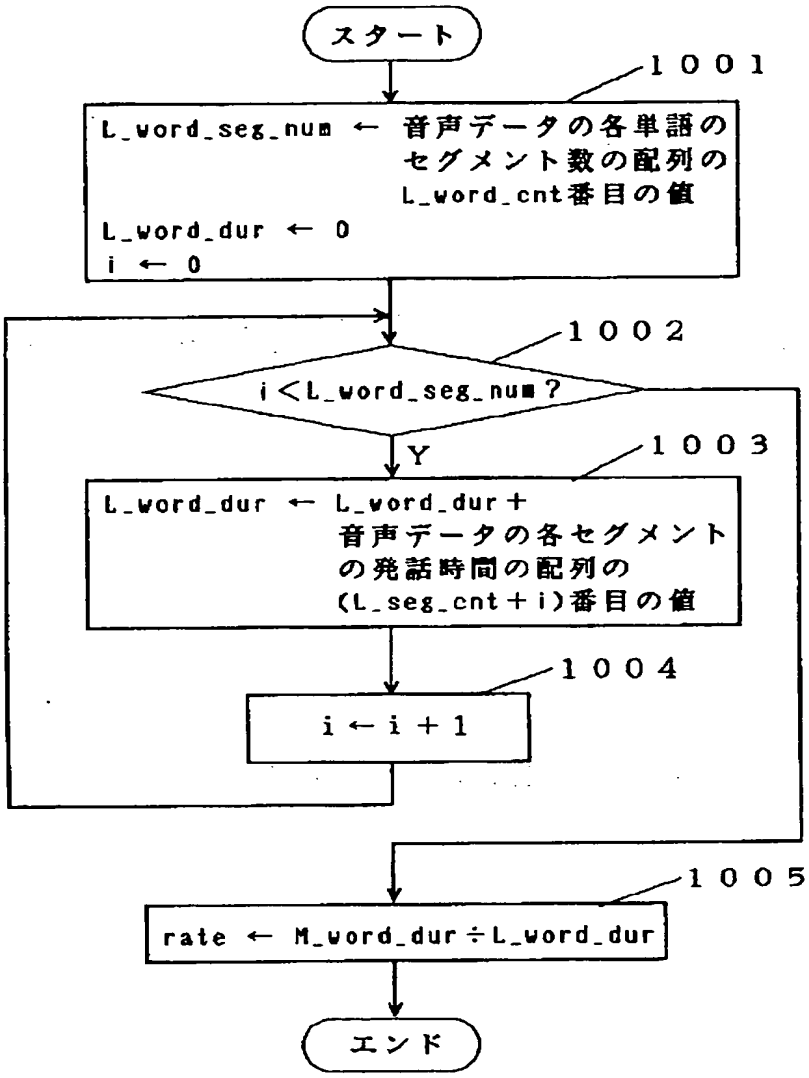
【図10】


スタート
801
X軸、Y軸表示
L_word_cnt←0　M_word_cnt←0
L_seg_cnt←0　M_seg_cnt←0
L_x[0]←0　M_x[0]←0
L_y[0]←0　M_y[0]←0
8011
L_word_num←音声データの単語数
802
L_word_cnt＜L_word_num？
N
エンド
Y
803
L_word_cnt番目の手本の単語の
イントネーション（ストレス）
グラフィック表示
805
学習者のL_word_cnt番目の単語に
対応するセグメントのデータを
手本に対応させて表示するための
倍率rateを計算
806
rateを利用して、L_word_cnt番目
の学習者の単語に対応する
イントネーション（ストレス）
のグラフィック表示
807
L_word_cnt ← L_word_cnt＋1

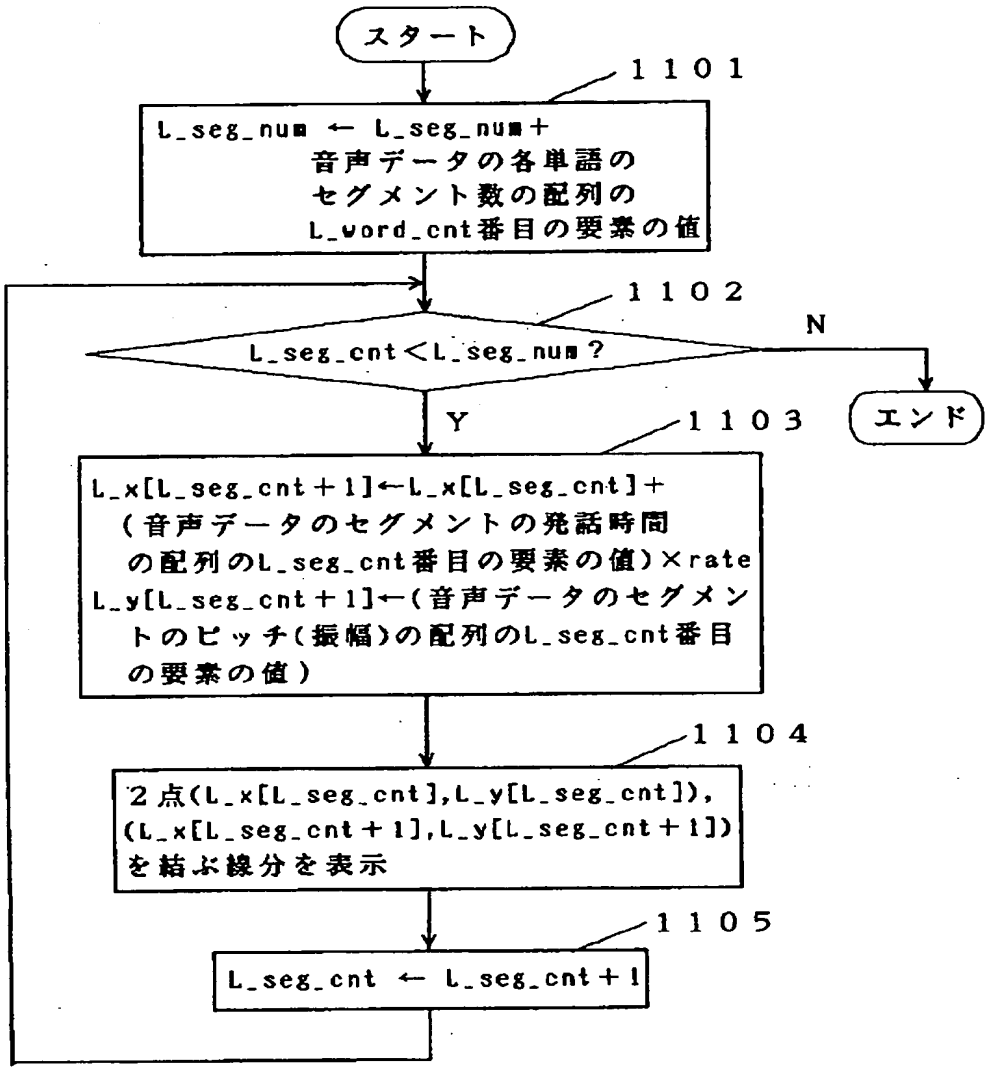
【図11】


スタート
901
M_word_dur←0
disp_word_flag←ON
902
M_seg_num ← M_seg_num
＋音声データの各単語の
セグメント数の配列の
M_word_cnt番目の値
B
903
M_seg_cnt＜M_seg_num？
N
エンド
Y
904
M_x[M_seg_cnt＋1]←L_x[L_seg_cnt]
＋手本音声データのセグメントの発話
時間の配列のM_seg_cnt番目の値
M_y[M_seg_cnt＋1]←手本音声データの
セグメントのピッチ(振幅)の配列の
M_seg_cnt番目の値
M_word_dur←M_word_dur
＋手本音声データのセグメントの発話
時間の配列のM_seg_cnt番目の値
905
2点(M_x[M_seg_cnt],M_y[M_seg_cnt]),
(M_x[M_seg_cnt＋1],M_y[M_seg_cnt＋1])
を結ぶ線分を表示
A

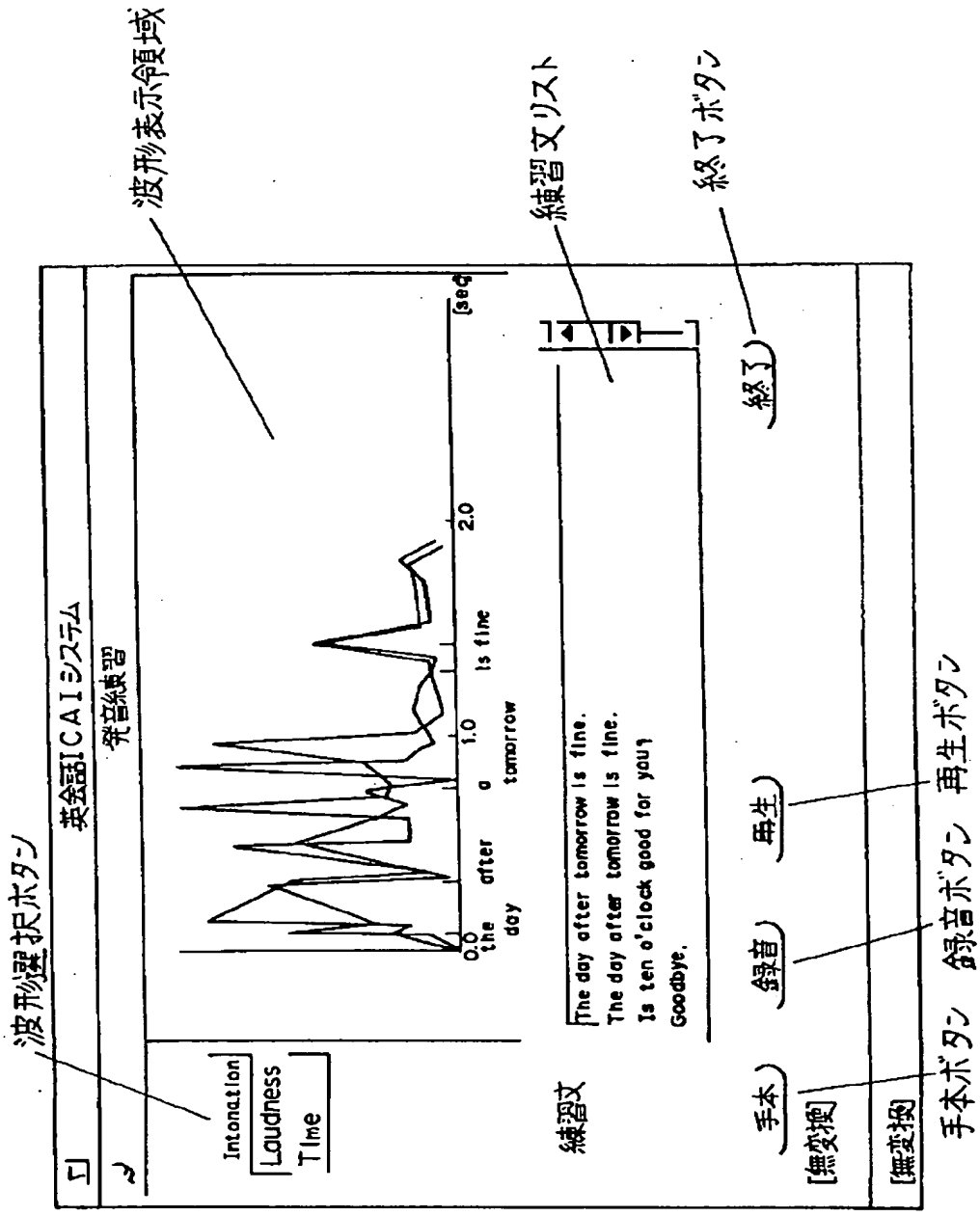
【図１３】


スタート
１００１
L_word_seg_num ← 音声データの各単語の
セグメント数の配列の
L_word_cnt番目の値
L_word_dur ← 0
i ← 0
１００２
i＜L_word_seg_num？
Y
１００３
L_word_dur ← L_word_dur＋
音声データの各セグメント
の発話時間の配列の
(L_seg_cnt＋i)番目の値
１００４
i ← i ＋ 1
１００５
rate ← M_word_dur÷L_word_dur
エンド

【図１４】


スタート
1101
L_seg_num ← L_seg_num＋
音声データの各単語の
セグメント数の配列の
L_word_cnt番目の要素の値
1102
L_seg_cnt＜L_seg_num？
N
エンド
Y
1103
L_x[L_seg_cnt＋1]←L_x[L_seg_cnt]＋
（音声データのセグメントの発話時間
の配列のL_seg_cnt番目の要素の値）×rate
L_y[L_seg_cnt＋1]←（音声データのセグメン
トのピッチ（振幅）の配列のL_seg_cnt番目
の要素の値）
1104
２点(L_x[L_seg_cnt],L_y[L_seg_cnt]),
(L_x[L_seg_cnt＋1],L_y[L_seg_cnt＋1])
を結ぶ線分を表示
1105
L_seg_cnt ← L_seg_cnt＋1

【図１５】



フロントページの続き

(72)発明者　田川　忠道
東京都港区虎ノ門１丁目７番12号　沖電気工業株式会社内

(72)発明者　加藤　正明
愛知県名古屋市千種区内山三丁目８番10号　株式会社沖テクノシステムズラボラトリ内